

2025 EECMS Summer Internship Application Form

Main Supervisor	Hoa Bui
Other supervisors (if applicable)	Vinesha Peiris
Project Title	Diversity selections for starting points of k-means algorithm
Student location(s) for the project	Building 314
Duration of project	Eight weeks
Project Description	<p>Clustering is an unsupervised partitioning technique that partitions data points into clusters based on similarity. It is a crucial method in data mining, used for detecting anomalies in network traffic and optimising resource allocation and logistics in military operations.</p> <p>The K-means algorithm is one of the most widely used and fast iterative clustering algorithms. However, it is known to be sensitive to initial starting conditions, including the initial cluster centroids and the order of instances. One strategy for selecting good starting points is to choose centroids with the highest dissimilarity measure. In this research, we aim to compare empirically different diversity selections of centroids together with two initialisation methods: random and Kaufman.</p>